



Conferencia Internacional
BIREDIAL - ISTEC
22-23-24-25 octubre 2018

Implementación de algoritmo para la
extracción de datos estructurados de
perfiles en Google Académico

Danny Murillo, Dalys Saavedra, Huriviades Calderón (Panamá)



Contenido

Introducción
Indicadores Bibliométricos
Plataformas
Alternativas
Minería de datos
Metodología
Resultados
Conclusión

Introducción

La idea de la web fue diseñar un espacio de trabajo colaborativo que facilitará el flujo de información.

Tim Berner-Lee, 1980

La sobreabundancia de información en Internet, es uno de los principales componentes de su éxito, sin embargo el tratamiento de esta exige una enorme cantidad de tiempo y energía a fin de cribar la calidad de los datos sumergidos en tan enorme repositorio.

Sánchez Carballido, 2011

Interés de las Universidades

Para las universidades se hace relevante obtener datos de fuentes que permitan identificar ¿quien publica? ¿dónde publica? ¿qué publica? Y cual es el impacto de lo que hace?

La medición es necesaria para toma de decisiones, reparto de fondos de investigación, otorgamiento de becas, justificación de las inversiones en I+D.

Periodic Table of Scientometric Indicators

C Total Citations		EC3 metrics		Basic Indicators Bibliometric Indicators H-index based Indicators												Webmetric Indicators (1.0)		Altmetric Indicators		Lnk Link	
h-i Index		P Number of Publications																			
IF Impact Factor	AF Article Factor	CS Citations	JCS Journal Citation Score	FCS Field Citation Score	FNCI Field-normalized citation indicators	NJI Normalized Journal Impact	JCS Journal Citation Score	RgC ResearchGate Citations	MASC Microsoft Academic Search Citations	GSC Google Scholar Citations	GSh-i Google Scholar Index	Lk Links	Fav Favorites	MR Mendeley Readers	AP Academia Publications	RGP ResearchGate Publications	WS Web of Science				
SJR Scimago Journal Rank	EF Eigenfactor	SNIP Source Normalized Impact Per Page	I3 Tripartite Impact Indicator	CI Cite Index	MCS Mean Citation Score	MNCS Mean Normalized Citation Score	MCRS Mean Citation Rank Score	MSNCS Mean Source Normalized Citation Score	MASP Microsoft Academic Search Papers	GSP Google Scholar Papers	Sub Subscribers	BM Blog Medium	TwM Twitter mentions	FacS Facebook Shares	ADV Academia Documents Viewed	RGV ResearchGate Viewed	Vw Views				
IPP Impact per Page	CPP Citation per paper	CPPex Citation per Paper with citation for in cited	ANCP Average Number of Citations per publication	TNCS Total number Average Number of Citations	RAI Relative Index to Average	RSI Relative Standardization Ratio	RCR Relative Citation Ratio	RDCP Relative Citation Citation Potential	JAR Journal Academic Rank	Com Comments	PuPC PubPeer Comments	NM News Mentions	WC Wikipedia Citations	FacC Facebook Comments	Afr Academia Followers	RGI ResearchGate Impact Ratio	Ck Checks				
%SC Self Citations	%Pnc Percentage of papers self-cited	PR Peer Review Ratio	LogZ Log score	IK Imputed Knowledge	TI Technological Impact	STP Scientific Talent Pool	NPJ Normalized Journal Impact	WCH Web of Science H-index	Rev Reviews	F1Re F1 Score Reviews	GoRev Goodreads Reviews	MoH Mendeley Holdings	ARev Amazon Reviews	Play YouTube Plays	Afg Academia Following	RGfr ResearchGate Following	FTV Full Text Views				
PT1 Papers in Top 1	PT10 Papers in Top 10	PT50 Papers in Top 50	HCP High Cited Papers	Q1 Papers in First Quartile	PWoS Highly Cited Papers in First Quartile	NHCP Normalized Highly Cited Papers in First Quartile	PTRJ Publication Top Ranked Journals	Exp Experts	Q&A Q&A Score	F1R F1 Score Reviews	GoRat Goodreads Rating	MoR Mendeley Rating	ARat Amazon Rating	PS Peer Score	OS Open Science	RGfg ResearchGate Following	AV Academia Views				
PCol Papers in Collaboration	%CoA Share of articles co-authored with another unit	NCoI Normalized Collaboration	ICol Interdisciplinary Collaboration	SL Scientific Leadership	EN Expertise Number	Exc Excellence	Sav Saves	ReR Research Recommendations	F1FFa F1 Score First	GoRea Goodreads Reviews	MoS Mendeley Score	RcCU ResearchGate Citations	RCU ResearchGate Citations	BoD Bibliometric Data	AA Academia Author	AAS Academia Author Score	DIL Digital Impact Link				

i10-i	g-i	a-i	h(2)-i	hg-i	q2-i	r-i	ar-i	k-i	f-i	m-i	m-q	Ch-i	Th-i	Dh-Ti	n-i	Mh-i
h5-i	Nh-i	SIs-i	Sih-Ti	Hw-i	Hm-i	Th-i	I10-i	v-i	e-i	m-i	Mh-i	RC-i	CC-i	Ch-i	CSs-i	π-i
h5-m	2gh-i	Rbhm-i	h2-l	h2-c	h2-u	h3-i	p-i	Hbar-i	Mhm-i	w-i	b-i	Gh-i	SPh-i	hint-i	Hrat-i	πv-i

Plataformas con indicadores Bibliométricos

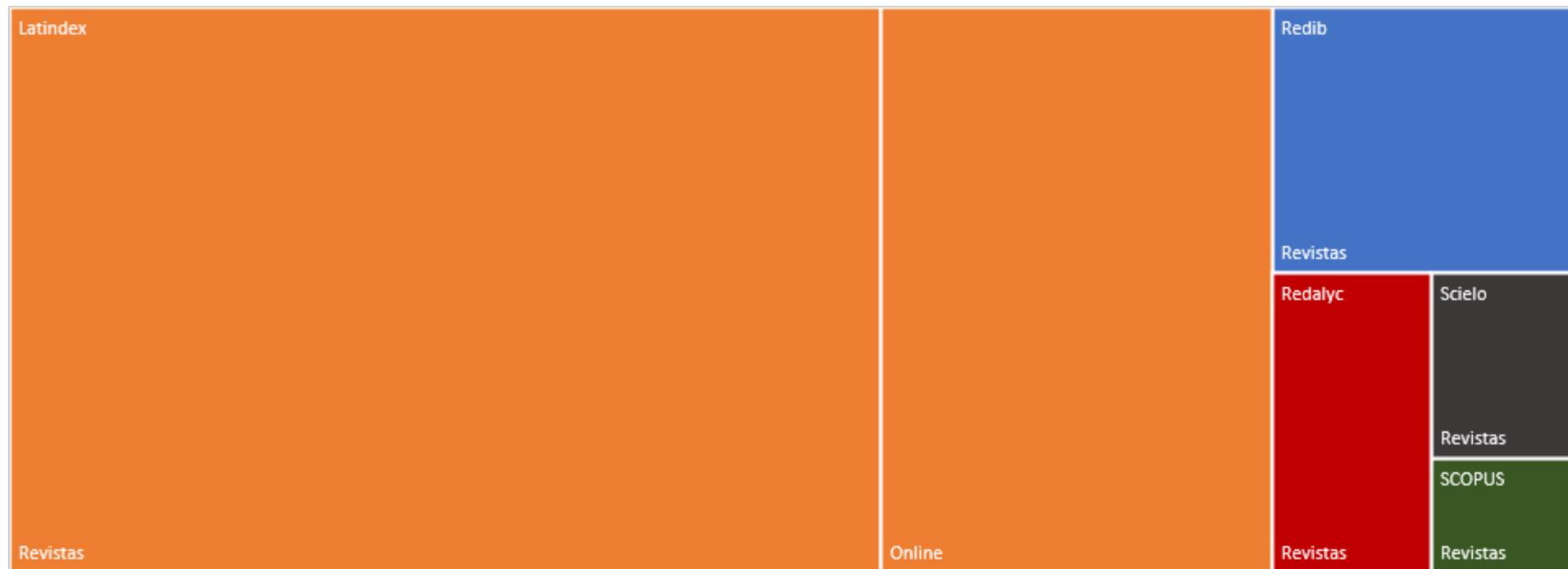


Scopus®

Números de Revistas en Indexadores Latinoamericanos

■ SCOPUS ■ Scielo ■ Latindex ■ Redalyc ■ Redib

País	Cantidad
Argentina	71
Chile	98
Colombia	69
Costa Rica	3
Ecuador	4
Mexico	132
Panamá	1
Peru	11
Puerto Rico	7
Uruguay	5
Venezuela	54



	Revistas
SCOPUS	1061
Scielo	737
Redalyc	1257
Latindex	12990
Redib	2176



Alternativas indicadores Bibliométricos

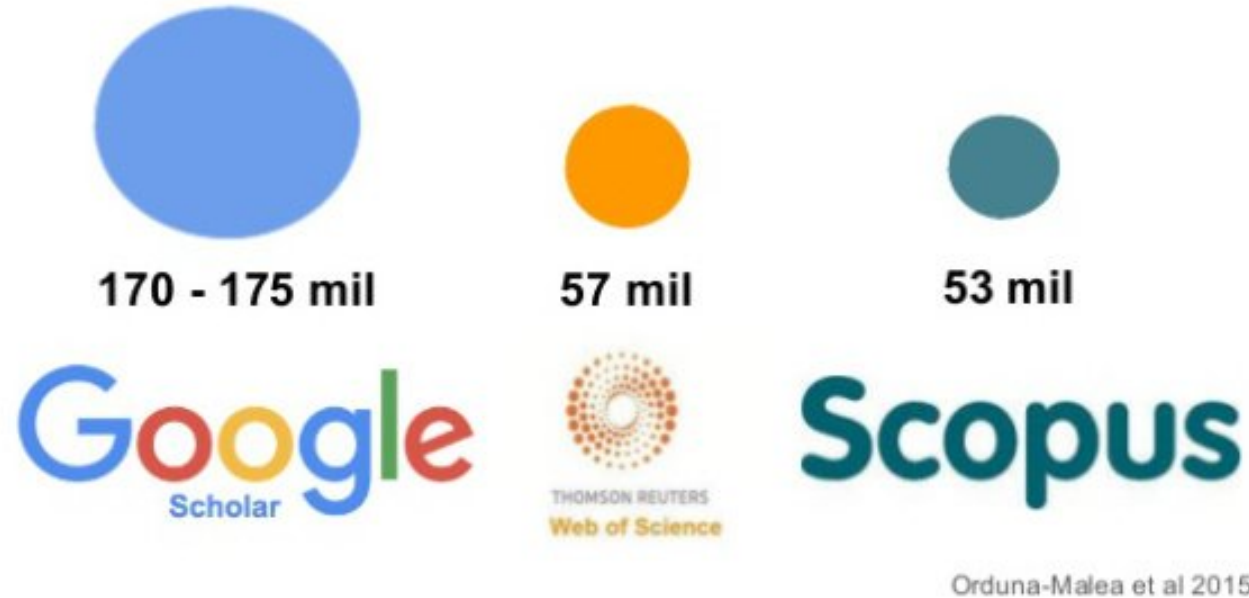
Google Académico

Perfiles
Publicaciones
Hindex

ResearchGate

Perfiles
Publicaciones
RGscore

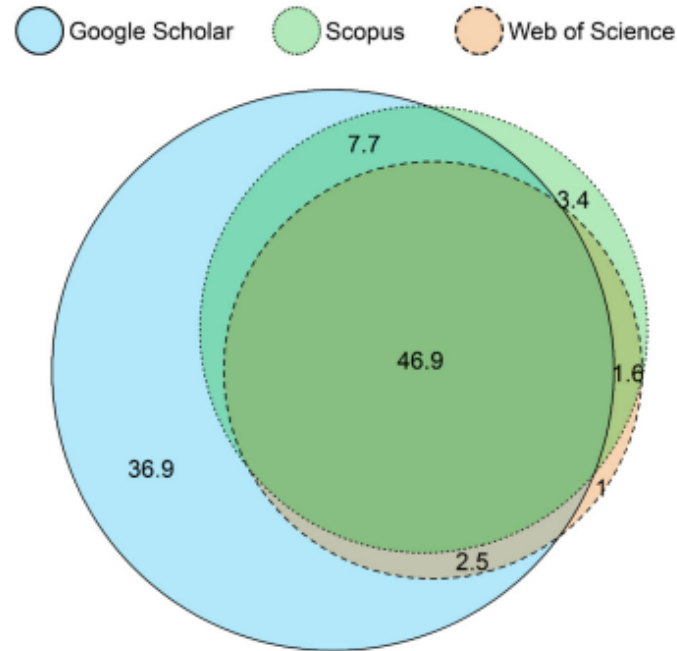
Does Size Matter?



Google Scholar, Web Of Science, y Scopus: una comparación sistemática de **citaciones** en 252 categorías de temas (Preprint)

Xavier Lasauca Cisa, 2017
Journal of Informetrics

Citas superpuestas (GS, Scopus, WoS)



Existe un 46,9% de todas las citas encontradas por las tres bases de datos. GS encontró la mayor cantidad de citas, incluida la mayoría de las citas encontradas por WoS y Scopus.

Las citas únicas y superpuestas en Google Scholar, Scopus, y Web of Science, $n = 2,448,055$ citas de todas las áreas temáticas.

Xavier Lasauca Cisa, 2017
Journal of Informetrics

Google Académico

Universidad Tecnológica de Panamá

[Más información](#)



Reinhardt Pinzon Adames

Centro de Investigaciones Hidraulicas Hidrotecnicas de la Universidad Tecnologica de ...
Dirección de correo verificada de utp.ac.pa

Citado por 1088



Elida de Obaldia

Universidad Tecnologica de Panama
Dirección de correo verificada de utp.ac.pa

Material Science Renewable Energy Diamond thin films Project based education

Citado por 995



Oscar M. Ramirez

Universidad Tecnologica de Panama
Dirección de correo verificada de utp.ac.pa

Structural and Earthquake Engi...

Citado por 485



Gilberto Axel Chang, PhD

Professor of Civil Engineering, Universidad Tecnológica de Panamá
Dirección de correo verificada de utp.ac.pa

Structural Engineering Seismic Engineering

Citado por 425



Héctor Montes Franceschi

Universidad Tecnológica de Panamá
Dirección de correo verificada de utp.ac.pa

Robotica Control

Citado por 391

Google Académico

Universidad Tecnológica de Panamá

[Más información](#)



Reinhardt Pinzon Adames

Centro de Investigaciones Hidraulicas Hidrotecnicas de la Universidad Tecnologica de ...
Dirección de correo verificada de utp.ac.pa

Citado por 1088



Elida de Obaldia

Universidad Tecnologica de Panama
Dirección de correo verificada de utp.ac.pa

[Material Science](#) [Renewable Energy](#) [Diamond thin films](#) [Project based education](#)

Citado por 995



Oscar M. Ramirez

Universidad Tecnologica de Panama
Dirección de correo verificada de utp.ac.pa

[Structural and Earthquake Engi...](#)

Citado por 485

ok, ¿Cómo extraigo los datos de
Google Scholar?



Reinhardt Pinzon Adames

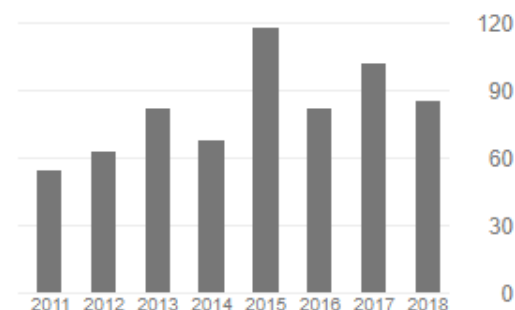
SEGUIR

Centro de Investigaciones Hidraulicas Hidrotecnicas de la [Universidad Tecnologica de Panama](#) (UTP)
Dirección de correo verificada de utp.ac.pa

CREAR MI PROPIO PERFIL

TÍTULO	CITADO POR	AÑO
Introduction to active thermochemical tables: Several "key" enthalpies of formation revisited B Ruscic, RE Pinzon, ML Morton, G von Laszewski, SJ Bittner, SG Nijsure, ... The Journal of Physical Chemistry A 108 (45), 9979-9997	364	2004
Active thermochemical tables: Accurate enthalpy of formation of hydroperoxyl radical, HO2 B Ruscic, RE Pinzon, ML Morton, NK Srinivasan, MC Su, JW Sutherland, ... The Journal of Physical Chemistry A 110 (21), 6592-6601	242	2006
Active Thermochemical Tables: thermochemistry for the 21st century B Ruscic, RE Pinzon, G Von Laszewski, D Kodeboyina, A Burcat, D Leahy, ... Journal of Physics: Conference Series 16 (1), 561	215	2005
Metadata in the collaboratory for multi-scale chemical science C Pancerella, J Hewson, W Koegler, D Leahy, M Lee, L Rahn, C Yang, ... International Conference on Dublin Core and Metadata Applications, 121-129	80	2003
Comparison of different three-site interaction potentials for liquid acetonitrile E Guardia, R Pinzón, J Casulleras, M Orozco, FJ Luque Molecular Simulation 26 (4), 287-306	49	2001
A collaborative informatics infrastructure for multi-scale science JD Myers, TC Allison, S Bittner, B Didier, M Frenklach, WH Green, YL Ho, ... Challenges of Large Applications in Distributed Environments, 2004. CLADE ...	38	2004

Citado por	VER TODO	
	Total	Desde 2013
Citas	1102	538
Índice h	10	7
Índice i10	10	5



Coautores



Branko Ruscic
Distinguished Fellow, Argonne N...



Gregor von Laszewski



Problemática

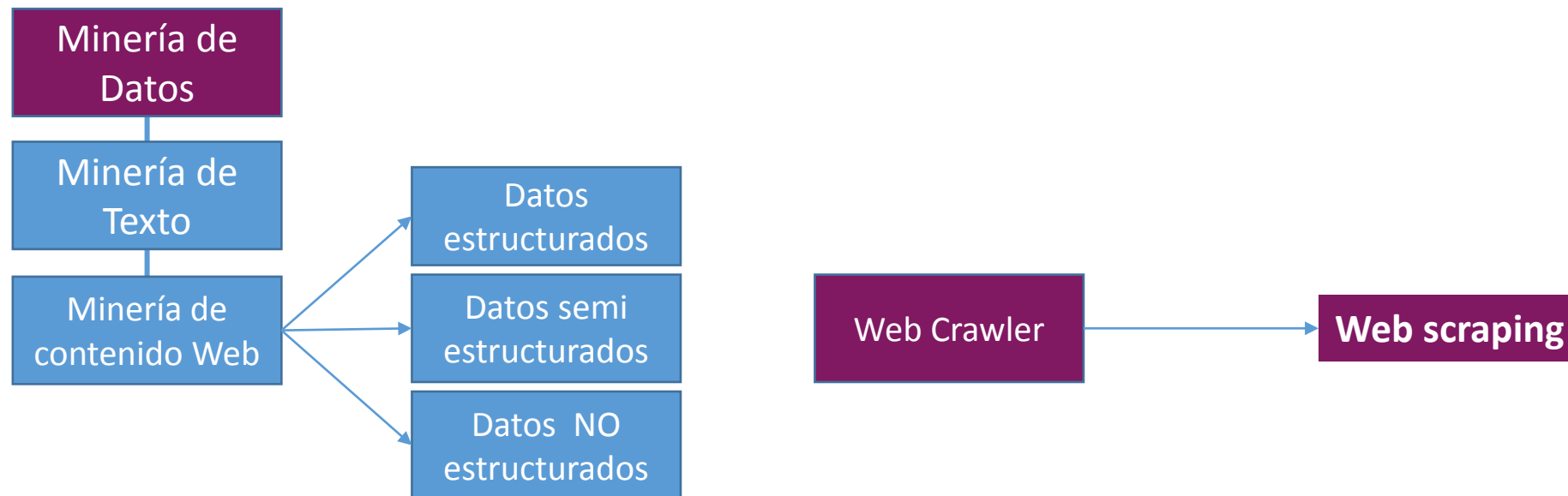
No existe una herramienta para extraer datos estructurados de Google Scholar (Perfiles y Publicaciones)

Objetivo

Evaluar la forma de automatizar el proceso extracción de los datos estructurados de los perfiles y publicaciones de una afiliación (universidad) en Google académico.

Minería de Datos

Es un campo de la estadística y las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos.



Web Scrapping

https://scholar.google.com/citations?view_op=view_org&org=4736061867397421563&hl=es&oi=io

Universidad Tecnológica de Panamá [Más información](#)



Reinhardt Pinzon Adames

Centro de Investigaciones Hidráulicas Hidrotecnicas de la Universidad Tecnologica de ...
Dirección de correo verificada de utp.ac.pa



Elida de Obaldia

Universidad Tecnologica de Panama
Dirección de correo verificada de utp.ac.pa
[Material Science](#) [Renewable Energy](#) [Diamond thin films](#) [Project based education](#)



Oscar M. Ramirez

Universidad Tecnologica de Panama
Dirección de correo verificada de utp.ac.pa
[Structural and Earthquake Engi...](#)



Gilberto Axel Chang, PhD

Professor of Civil Engineering, Universidad Tecnológica de Panamá
Dirección de correo verificada de utp.ac.pa
[Structural Engineering](#) [Seismic Engineering](#)



Héctor Montes Franceschi

Universidad Tecnológica de Panamá
Dirección de correo verificada de utp.ac.pa
[Robotica](#) [Control](#)

Web Scrapping

Universidad Tecnológica de Panamá Más información



Reinhardt Pinzon Adames

Centro de Investigaciones Hidráulicas Hidroeléctricas de la Universidad Tecnológica de ...
Dirección de correo verificada de utp.ac.pa



Elida de Obaldia

Universidad Tecnológica de Panamá
Dirección de correo verificada de utp.ac.pa
Material Science Renewable Energy Diamond thin films Project based education



Oscar M. Ramirez

Universidad Tecnológica de Panamá
Dirección de correo verificada de utp.ac.pa
Structural and Earthquake Engi...



Gilberto Axel Chang, PhD

Professor of Civil Engineering, Universidad Tecnológica de Panamá
Dirección de correo verificada de utp.ac.pa
Structural Engineering Seismic Engineering



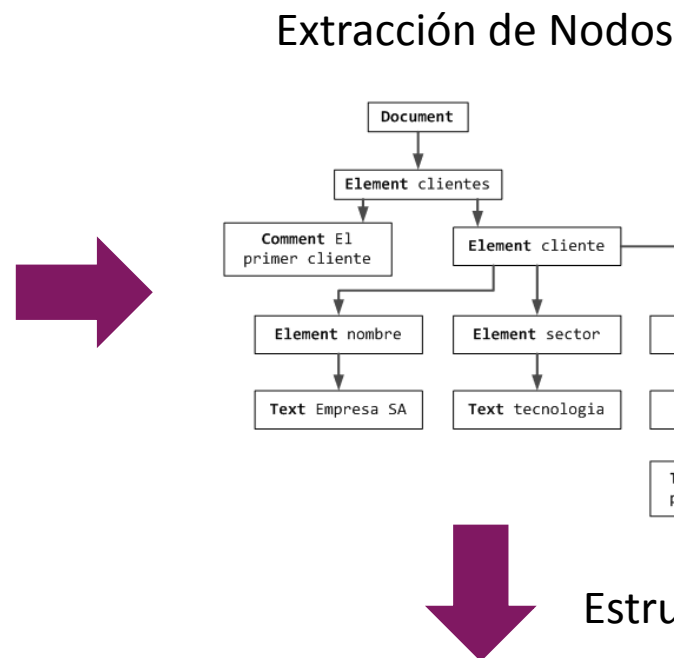
Héctor Montes Franceschi

Universidad Tecnológica de Panamá
Dirección de correo verificada de utp.ac.pa
Robotica Control

```
empty{display:none}</style><div id="gs_alrt_w" role="alert"><div id="gs_alrt_p"><form><input
id="gs_alrt_ss"></form><form action="" method="post" id="gs_alrt" class="gs_alrt"><span
id="gs_alrt_m"></span><span id="gs_alrt_h"></span><a id="gs_alrt_l" href="javascript:void(0)"
onclick="document.getElementById('gs_alrt').submit()"></a></form></div><div id="gs_ab"
class=""><div id="gs_ab_ico" class="gs_btnGSL"><span class="gs_ico"></span></div><div
id="gs_ab_ttl"><div class="gs_ab_md"Perfiles</div></div><div id="gs_ab_md"><div
class="gs_ab_md"></div></div><div id="gs_ab_btns"><a href="/citations?hl=es" class="gs_btnPRO
gs_in_ib gs_nph gs_nta"><span class="gs_ico"></span><span class="gs_lbl">Mi perfil</span></a><a
href="/scholar?scilib=1&hl=es" class="gs_btnL gs_in_ib gs_nph gs_nta"><span class="gs_ico">
</span><span class="gs_lbl">Mi biblioteca</span></a></div></div><div id="gs_bdy"><div
id="gs_bdy_sb" role="navigation"><div id="gs_bdy_sb_in"></div></div><div id="gs_bdy_ccl"
role="main"><div id="gsc_sa_ccl"><h2 class="gsc_authors_header">Universidad Tecnológica de Panamá
<a id="gsc_org_help_label" href="javascript:void(0)">Más información</a></h2><div
id="gsc_org_help">En esta página se indican perfiles públicos para Universidad Tecnológica de Panamá
y verifica una dirección de correo electrónico de esta institución. <a id="gsc_org_help_hide"
href="javascript:void(0)">ocultar</a></div><div class="gsc_lusr gs_scl"><span class="gs_rimg
gsc_pp_sm gsc_lusr_photo"></span><div
class="gsc_oai"><h3 class="gsc_oai_name"><a href="/citations?
user=1TICxmUAAAAJ&hl=es">Reinhardt Pinzon Adames</a></h3><div class="gsc_oai_aff">Centro de
Investigaciones Hidráulicas Hidroeléctricas de la Universidad Tecnológica de ...</div><div
class="gsc_oai_eml">Dirección de correo verificada de utp.ac.pa</div><div
class="gsc_oai_cby">Citado por 1102</div><div class="gsc_oai_int"></div></div><div
class="gsc_lusr gs_scl"><span class="gs_rimg gsc_pp_sm gsc_lusr_photo"></span><div
class="gsc_oai"><h3 class="gsc_oai_name"><a href="/citations?user=18gpxI4AAAAJ&hl=es">Elida
de Obaldia</a></h3><div class="gsc_oai_aff">Universidad Tecnológica de Panamá</div><div
class="gsc_oai_eml">Dirección de correo verificada de utp.ac.pa</div><div
class="gsc_oai_cby">Citado por 1005</div><div class="gsc_oai_int"><a class="gsc_oai_one_int"
href="/citations?view_op=search_authors&hl=es&mauthors=label:material_science">Material
Science</a> <a class="gsc_oai_one_int" href="/citations?
view_op=search_authors&hl=es&mauthors=label:renewable_energy">Renewable Energy</a> <a
class="gsc_oai_one_int" href="/citations?
view_op=search_authors&hl=es&mauthors=label:diamond_thin_films">Diamond thin films</a> <a
```

Web Scraping

Análisis datos en el DOM

[illegible]

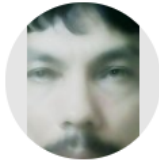
	Nombre	citas	word_key	url_user	Id_user
1	Alberto Gago	28003	Particle Physics, Neutrino Physics & Astrophysics	/citations?user=qRf0l-8AAAAJ&hl=es&oe=ASCII	qRf0l-8AAAAJ
2	Jose Bazo	7852	Astroparticle physics, Particle physics, High energy ...	/citations?user=axR80WoAAAAJ&hl=es&oe=ASCII	axR80WoAAAAJ
3	Alex Coad	5377	Economics of Innovation, Industrial Economics, Firm ...	/citations?user=hyoV0CmIAAAJ&hl=es&oe=ASCII	hyoV0CmIAAAJ
4	Norma Salinas	2445	Tropical ecology, Ecosystem dynamics, Plant anatom...	/citations?user=CRtKj-4AAAAJ&hl=es&oe=ASCII	CRtKj-4AAAAJ
5	Adolfo Figueroa	2361	teoria economica, desarrollo economico, creciment...	/citations?user=ndrlqQIAAAJ&hl=es&oe=ASCII	ndrlqQIAAAJ
6	Martin Tanaka	2272	ciencia politica	/citations?user=ajtldvYAAAAJ&hl=es&oe=ASCII	ajtldvYAAAAJ
7	Dora Isabel Herrera Paredes de Del Aguila	2255	Motivation, Future Time Perspective, Self Determinat...	/citations?user=9lxBrMIAAAAJ&hl=es&oe=ASCII	9lxBrMIAAAAJ
8	Norma Fuller-Osores	2150	Cultural diversity, Gender, Development and tourism, ...	/citations?user=QmlspHAAAAAJ&hl=es&oe=ASCII	QmlspHAAAAAJ
9	PIERO MOROSINI	2105	leadership, mindfulness, mergers and acquisitions, in...	/citations?user=IT932tQAAAAJ&hl=es&oe=ASCII	IT932tQAAAAJ
10	Roxana Barrantes	2085	regulación, infraestructura, TIC, desarrollo económic...	/citations?user=NivZRyYAAAAJ&hl=es&oe=ASCII	NivZRyYAAAAJ
11	IAN VÁZQUEZ ROWE	1880	Carbon footprint, Life Cycle Assessment, Fisheries a...	/citations?user=GxClvRkIAAAJ&hl=es&oe=ASCII	GxClvRkIAAAJ
12	Fernando Gilberto Torres Garcia	1788		/citations?user=v2mDNg0AAAAJ&hl=es&oe=ASCII	v2mDNg0AAAAJ
13	Daniel Racoceanu	1770	BioMedical Image and Data Computing	/citations?user=2eBRLjQAAAAJ&hl=es&oe=ASCII	2eBRLjQAAAAJ
14	Gregory J. Scott	1705		/citations?user=Fqf-hSMAAAJ&hl=es&oe=ASCII	Fqf-hSMAAAJ
15	Luis Jaime Castillo	1692	Arqueología Andina, Mochica, Ideología y Poder, Dro...	/citations?user=pYeFuTAAAAJ&hl=es&oe=ASCII	pYeFuTAAAAJ

¿Cómo Implementar web Scraping?

- A través de Softwares
 - Scraper – Plugin Chrome
 - Fminer - Escritorio
 - Import.io - Online
- A través de Implementaciones desarrolladas con algún lenguaje (Phyton, R)
 - **Función en R “GScholarScraper”** : función en R creada en el 2012 por Kay Cichini ,extrae detalles de las publicaciones en GS.
 - **Paquete en R “scholar”**: creado por James Keirsted en el 2015, proporciona funciones para extraer datos de GS.

Web Scraping utilizando diferentes Software

Promedio de perfiles (55) , Publicaciones (200)



Reinhardt Pinzon Adames

Centro de Investigaciones Hidraulicas Hidrotecnicas de la [Universidad Tecnologica de Panama](#) (UTP)
Dirección de correo verificada de utp.ac.pa

SEGUIR

TÍTULO	CITADO POR	AÑO
Introduction to active thermochemical tables: Several "key" enthalpies of formation revisited B Ruscic, RE Pinzon, ML Morton, G von Laszewski, SJ Bittner, SG Nijssure, ... The Journal of Physical Chemistry A 108 (45), 9979-9997	364	2004
Active thermochemical tables: Accurate enthalpy of formation of hydroperoxyl radical, HO2 B Ruscic, RE Pinzon, ML Morton, NK Srinivasan, MC Su, JW Sutherland, ... The Journal of Physical Chemistry A 110 (21), 6592-6601	242	2006
Active Thermochemical Tables: thermochemistry for the 21st century B Ruscic, RE Pinzon, G Von Laszewski, D Kodeboyina, A Burcat, D Leahy, ... Journal of Physics: Conference Series 16 (1), 561	215	2005

Universidad	TIEMPO POR MÉTODO WEB SCRAPING			
	Perfiles / Publicaciones (minutos)			
	Copiar / Pegar	Local Browser	Local Software	Online
UFM	8 / 130	2 / 35	3 / 50	2 / 35
ESPOL	42 / 354	9 / 95	14 / 140	9 / 94
FURB	24 / 445	5 / 122	8 / 179	5 / 120
UTP	50 / 482	11/129	17 / 189	10 / 127
UH	51 / 920	11 / 245	17 / 363	10 / 244
Promedio	35 / 466	8 / 125	12 / 184	7 / 124

Scraper

FMiner

Import.io

Web Scraping con paquete Scholar de R



Reinhardt Pinzon Adames

SEGUIR

Centro de Investigaciones Hidraulicas Hidrotecnicas de la [Universidad Tecnologica de Panama](#) (UTP)
Dirección de correo verificada de utp.ac.pa

TÍTULO	CITADO POR	AÑO
Introduction to active thermochemical tables: Several "key" enthalpies of formation revisited B Ruscic, RE Pinzon, ML Morton, G von Laszewski, SJ Blittner, SG Nijssure, ... The Journal of Physical Chemistry A 108 (45), 9979-9997	364	2004
Active thermochemical tables: Accurate enthalpy of formation of hydroperoxyl radical, HO2 B Ruscic, RE Pinzon, ML Morton, NK Srinivasan, MC Su, JW Sutherland, ... The Journal of Physical Chemistry A 110 (21), 6592-6601	242	2006
Active Thermochemical Tables: thermochemistry for the 21st century B Ruscic, RE Pinzon, G Von Laszewski, D Kodeboyina, A Burcat, D Leahy, ... Journal of Physics: Conference Series 16 (1), 561	215	2005

Universidad	País	Perfiles		Publicaciones		Tiempo
		manual	extraídos	manual	extraídas	
UFM	Guatemala	14	14	393	393	3
ESPOL	Ecuador	67	67	1061	1061	3
FURB	Brasil	38	38	1360	1360	4
UTP	Panamá	77	77	1434	1434	4
UH	Cuba	79	79	2758	2758	5

Web Scraping con paquete Scholar

Universidad	#Perfiles correctos	Con paquete Scholar	
		#Perfiles	Tiempo (minutos)
UDELAR	182	182	2
UCR	230	230	2
UFC	119	119	2
UDEA	383	383	3
UCHILE	566	137	5
UNAM	1329	138	11
OSAKAU	460	140	4
UED	1471	147	14
UPV	794	387	11
UILLINOIS	2555	2250	62
	8364	4388	122

Esquema del Algoritmo en R desarrollado

Cargar librería en R

Cargar Funciones en R

Leer url de afiliación de la universidad en GA

Escanear página de afiliación



Leer nodos html con datos

[illegible]

Extraer indicadores métricos de perfiles

[illegible]

Escanear p gi
perfiles



Leer url de cada perfil

Extraer url de perfiles en formato de tabla

Bibliography		Accession Number
1. Bhatia S, Bhatia J	6121	Accession Number: 61210001
2. Bhatia S, Bhatia J	6122	Accession Number: 61220001
3. Bhatia S, Bhatia J	6123	Accession Number: 61230001
4. Bhatia S, Bhatia J	6124	Accession Number: 61240001
5. Bhatia S, Bhatia J	6125	Accession Number: 61250001
6. Bhatia S, Bhatia J	6126	Accession Number: 61260001
7. Bhatia S, Bhatia J	6127	Accession Number: 61270001
8. Bhatia S, Bhatia J	6128	Accession Number: 61280001
9. Bhatia S, Bhatia J	6129	Accession Number: 61290001
10. Bhatia S, Bhatia J	6130	Accession Number: 61300001
11. Bhatia S, Bhatia J	6131	Accession Number: 61310001
12. Bhatia S, Bhatia J	6132	Accession Number: 61320001
13. Bhatia S, Bhatia J	6133	Accession Number: 61330001
14. Bhatia S, Bhatia J	6134	Accession Number: 61340001
15. Bhatia S, Bhatia J	6135	Accession Number: 61350001
16. Bhatia S, Bhatia J	6136	Accession Number: 61360001
17. Bhatia S, Bhatia J	6137	Accession Number: 61370001
18. Bhatia S, Bhatia J	6138	Accession Number: 61380001
19. Bhatia S, Bhatia J	6139	Accession Number: 61390001
20. Bhatia S, Bhatia J	6140	Accession Number: 61400001
21. Bhatia S, Bhatia J	6141	Accession Number: 61410001
22. Bhatia S, Bhatia J	6142	Accession Number: 61420001
23. Bhatia S, Bhatia J	6143	Accession Number: 61430001
24. Bhatia S, Bhatia J	6144	Accession Number: 61440001
25. Bhatia S, Bhatia J	6145	Accession Number: 61450001
26. Bhatia S, Bhatia J	6146	Accession Number: 61460001
27. Bhatia S, Bhatia J	6147	Accession Number: 61470001
28. Bhatia S, Bhatia J	6148	Accession Number: 61480001
29. Bhatia S, Bhatia J	6149	Accession Number: 61490001
30. Bhatia S, Bhatia J	6150	Accession Number: 61500001
31. Bhatia S, Bhatia J	6151	Accession Number: 61510001
32. Bhatia S, Bhatia J	6152	Accession Number: 61520001
33. Bhatia S, Bhatia J	6153	Accession Number: 61530001
34. Bhatia S, Bhatia J	6154	Accession Number: 61540001
35. Bhatia S, Bhatia J	6155	Accession Number: 61550001
36. Bhatia S, Bhatia J	6156	Accession Number: 61560001
37. Bhatia S, Bhatia J	6157	Accession Number: 61570001
38. Bhatia S, Bhatia J	6158	Accession Number: 61580001
39. Bhatia S, Bhatia J	6159	Accession Number: 61590001
40. Bhatia S, Bhatia J	6160	Accession Number: 61600001
41. Bhatia S, Bhatia J	6161	Accession Number: 61610001
42. Bhatia S, Bhatia J	6162	Accession Number: 61620001
43. Bhatia S, Bhatia J	6163	Accession Number: 61630001
44. Bhatia S, Bhatia J	6164	Accession Number: 61640001
45. Bhatia S, Bhatia J	6165	Accession Number: 61650001
46. Bhatia S, Bhatia J	6166	Accession Number: 61660001
47. Bhatia S, Bhatia J	6167	Accession Number: 61670001
48. Bhatia S, Bhatia J	6168	Accession Number: 61680001
49. Bhatia S, Bhatia J	6169	Accession Number: 61690001
50. Bhatia S, Bhatia J	6170	Accession Number: 61700001
51. Bhatia S, Bhatia J	6171	Accession Number: 61710001
52. Bhatia S, Bhatia J	6172	Accession Number: 61720001
53. Bhatia S, Bhatia J	6173	Accession Number: 61730001
54. Bhatia S, Bhatia J	6174	Accession Number: 61740001
55. Bhatia S, Bhatia J	6175	Accession Number: 61750001
56. Bhatia S, Bhatia J	6176	Accession Number: 61760001
57. Bhatia S, Bhatia J	6177	Accession Number: 61770001
58. Bhatia S, Bhatia J	6178	Accession Number: 61780001
59. Bhatia S, Bhatia J	6179	Accession Number: 61790001
60. Bhatia S, Bhatia J	6180	Accession Number: 61800001
61. Bhatia S, Bhatia J	6181	Accession Number: 61810001
62. Bhatia S, Bhatia J	6182	Accession Number: 61820001
63. Bhatia S, Bhatia J	6183	Accession Number: 61830001
64. Bhatia S, Bhatia J	6184	Accession Number: 61840001
65. Bhatia S, Bhatia J	6185	Accession Number: 61850001
66. Bhatia S, Bhatia J	6186	Accession Number: 61860001
67. Bhatia S, Bhatia J	6187	Accession Number: 61870001
68. Bhatia S, Bhatia J	6188	Accession Number: 61880001
69. Bhatia S, Bhatia J	6189	Accession Number: 61890001
70. Bhatia S, Bhatia J	6190	Accession Number: 61900001
71. Bhatia S, Bhatia J	6191	Accession Number: 61910001
72. Bhatia S, Bhatia J	6192	Accession Number: 61920001
73. Bhatia S, Bhatia J	6193	Accession Number: 61930001
74. Bhatia S, Bhatia J	6194	Accession Number: 61940001
75. Bhatia S, Bhatia J	6195	Accession Number: 61950001
76. Bhatia S, Bhatia J	6196	Accession Number: 61960001
77. Bhatia S, Bhatia J	6197	Accession Number: 61970001
78. Bhatia S, Bhatia J	6198	Accession Number: 61980001
79. Bhatia S, Bhatia J	6199	Accession Number: 61990001
80. Bhatia S, Bhatia J	6200	Accession Number: 62000001
81. Bhatia S, Bhatia J	6201	Accession Number: 62010001
82. Bhatia S, Bhatia J	6202	Accession Number: 62020001
83. Bhatia S, Bhatia J	6203	Accession Number: 62030001
84. Bhatia S, Bhatia J	6204	Accession Number: 62040001
85. Bhatia S, Bhatia J	6205	Accession Number: 62050001
86. Bhatia S, Bhatia J	6206	Accession Number: 62060001
87. Bhatia S, Bhatia J	6207	Accession Number: 62070001
88. Bhatia S, Bhatia J	6208	Accession Number: 62080001
89. Bhatia S, Bhatia J	6209	Accession Number: 62090001
90. Bhatia S, Bhatia J	6210	Accession Number: 62100001
91. Bhatia S, Bhatia J	6211	Accession Number: 62110001
92. Bhatia S, Bhatia J	6212	Accession Number: 62120001
93. Bhatia S, Bhatia J	6213	Accession Number: 62130001
94. Bhatia S, Bhatia J	6214	Accession Number: 62140001
95. Bhatia S, Bhatia J	6215	Accession Number: 62150001
96. Bhatia S, Bhatia J	6216	Accession Number: 62160001
97. Bhatia S, Bhatia J	6217	Accession Number: 62170001
98. Bhatia S, Bhatia J	6218	Accession Number: 62180001
99. Bhatia S, Bhatia J	6219	Accession Number: 62190001
100. Bhatia S, Bhatia J	6220	Accession Number: 62200001

Leer datos de perfiles
de la tabla

Leer url de cada perfil

Escanear página de publicaciones



Extraer datos de publicaciones de perfil

[illegible]

Repetir

Gráfica de # de citas y publicaciones



Unir datos de las publicaciones por perfil

Index	Variable	Mean	Standard deviation	Minimum	Maximum
1	Age	39.23	10.23	20	60
2	Gender	0.49	0.50	0	1
3	Marital status	0.51	0.50	0	1
4	Education	12.51	1.50	9	16
5	Income	1.25	0.40	0.50	2.00
6	Health status	0.51	0.50	0	1
7	Life satisfaction	4.51	0.50	4	5
8	Life satisfaction	4.51	0.50	4	5
9	Life satisfaction	4.51	0.50	4	5
10	Life satisfaction	4.51	0.50	4	5
11	Life satisfaction	4.51	0.50	4	5
12	Life satisfaction	4.51	0.50	4	5
13	Life satisfaction	4.51	0.50	4	5
14	Life satisfaction	4.51	0.50	4	5
15	Life satisfaction	4.51	0.50	4	5
16	Life satisfaction	4.51	0.50	4	5
17	Life satisfaction	4.51	0.50	4	5
18	Life satisfaction	4.51	0.50	4	5
19	Life satisfaction	4.51	0.50	4	5
20	Life satisfaction	4.51	0.50	4	5
21	Life satisfaction	4.51	0.50	4	5
22	Life satisfaction	4.51	0.50	4	5
23	Life satisfaction	4.51	0.50	4	5
24	Life satisfaction	4.51	0.50	4	5
25	Life satisfaction	4.51	0.50	4	5
26	Life satisfaction	4.51	0.50	4	5
27	Life satisfaction	4.51	0.50	4	5
28	Life satisfaction	4.51	0.50	4	5
29	Life satisfaction	4.51	0.50	4	5
30	Life satisfaction	4.51	0.50	4	5
31	Life satisfaction	4.51	0.50	4	5
32	Life satisfaction	4.51	0.50	4	5
33	Life satisfaction	4.51	0.50	4	5
34	Life satisfaction	4.51	0.50	4	5
35	Life satisfaction	4.51	0.50	4	5
36	Life satisfaction	4.51	0.50	4	5
37	Life satisfaction	4.51	0.50	4	5
38	Life satisfaction	4.51	0.50	4	5
39	Life satisfaction	4.51	0.50	4	5
40	Life satisfaction	4.51	0.50	4	5
41	Life satisfaction	4.51	0.50	4	5
42	Life satisfaction	4.51	0.50	4	5
43	Life satisfaction	4.51	0.50	4	5
44	Life satisfaction	4.51	0.50	4	5
45	Life satisfaction	4.51	0.50	4	5
46	Life satisfaction	4.51	0.50	4	5
47	Life satisfaction	4.51	0.50	4	5
48	Life satisfaction	4.51	0.50	4	5
49	Life satisfaction	4.51	0.50	4	5
50	Life satisfaction	4.51	0.50	4	5

Agrupar # de citas y publicaciones por perfil

Index	Text	Index	Text	Index	Text
1	1. The first step in the process is to identify the problem.	101	101. The first step in the process is to identify the problem.	201	201. The first step in the process is to identify the problem.
2	2. The second step is to analyze the problem.	102	102. The second step is to analyze the problem.	202	202. The second step is to analyze the problem.
3	3. The third step is to develop a plan.	103	103. The third step is to develop a plan.	203	203. The third step is to develop a plan.
4	4. The fourth step is to implement the plan.	104	104. The fourth step is to implement the plan.	204	204. The fourth step is to implement the plan.
5	5. The fifth step is to evaluate the results.	105	105. The fifth step is to evaluate the results.	205	205. The fifth step is to evaluate the results.
6	6. The sixth step is to report the findings.	106	106. The sixth step is to report the findings.	206	206. The sixth step is to report the findings.
7	7. The seventh step is to disseminate the information.	107	107. The seventh step is to disseminate the information.	207	207. The seventh step is to disseminate the information.
8	8. The eighth step is to monitor the progress.	108	108. The eighth step is to monitor the progress.	208	208. The eighth step is to monitor the progress.
9	9. The ninth step is to review the process.	109	109. The ninth step is to review the process.	209	209. The ninth step is to review the process.
10	10. The tenth step is to conclude the project.	110	110. The tenth step is to conclude the project.	210	210. The tenth step is to conclude the project.

Web Scrapping con y sin paquete Scholar

		Con paquete Scholar		Sin paquete Scholar	
Universidad	#Perfiles correctos	#Perfiles	Tiempo (minutos)	#Perfiles	Tiempo (minutos)
UDELAR	182	182	2	182	1
UCR	230	230	2	230	1
UFC	119	119	2	119	1
UDEA	383	383	3	383	1
UCHILE	566	137	5	566	2
UNAM	1329	138	11	1329	5
OSAKAU	460	140	4	460	1
UED	1471	147	14	1471	5
UPV	794	387	11	794	3
UILLINOIS	2555	2250	62	2555	12
	8364	4388	122	8364	38

Prueba de algoritmo Scholar en perfil de 30 Universidades de Centroamérica



	Pais	Universidad	Siglas	perfiles	publicaciones	citaciones	hindex	anos
1	Puerto Rico	Universidad de Puerto Rico	UPR	296	13541	257257	9.2	162
2	Costa Rica	Universidad de Costa Rica	UCR	245	11110	133274	7.3	196
3	Jamaica	University of the West Indies	UWI	337	9896	173503	7.3	114
4	Trinidad and Tobago	University of the West Indies at St Augustine	UWA	205	7877	83120	7.1	267
5	Cuba	Universidad de la Habana	UHC	119	3453	22186	4.8	64
6	Costa Rica	Centro Agronómico Tropical de Investigación y Enseñanza	CAT	56	3846	33801	6.7	91
7	Costa Rica	Universidad Nacional de Costa Rica	UNA	115	5324	26263	4.7	195
8	Cuba	Universidad Central Marta Abreu de la Villas	UCMA	61	1793	7119	3.9	106
9	Costa Rica	Tecnológico de Costa Rica	TEC	264	6918	22855	2.6	116
10	Guatemala	Universidad Francisco Marroquín	UFM	13	397	2472	4.8	184
11	Guatemala	Universidad del Valle de Guatemala	UVG	24	753	12053	7.5	65
12	Puerto Rico	Ponce School of Medicine	PONCE	12	1273	19122	13.5	97
13	Panamá	Universidad Tecnológica de Panamá	UTP	87	1592	6713	2.8	39
14	Costa Rica	INCAE Business School	INCAE	9	392	2943	7.3	155
15	Cuba	Universidad de Oriente Santiago de Cuba	UOSC	37	580	1458	1.9	31
16	Cuba	Universidad de Pinar del Rio	UPIR	12	517	2038	4.3	81
17	República Dominicana	Pontificia Universidad Católica Madre y Maestra	PUCMM	42	480	2442	2.5	106
18	Cuba	Instituto Superior Politécnico José Antonio Echeverría	ISP	87	1888	5825	2.6	316
19	Costa Rica	Universidad Centroamericana José Simeón Cañas	UCC	7	130	246	2.4	17
20	Jamaica	University of Technology Jamaica	UTJ	12	178	1022	4.1	35
21	Puerto Rico	Universidad Politécnica de Puerto Rico	PUPR	8	159	1011	4.4	28
22	Costa Rica	Universidad Estatal a Distancia Costa Rica	UEaD	26	600	654	1.4	54
23	Cuba	Universidad de Ciencias Informáticas	UCI	74	1841	1197	1.3	45
24	Honduras	Escuela Agrícola Panamericana Zamorano	EAPZ	15	1273	3108	12.1	97
25	El Salvador	Universidad Don Bosco El Salvador	UDS	20	202	226	3.2	36
26	Cuba	Universidad de Camaguey	UCAC	47	1636	3640	3.1	48
27	Cuba	Instituto Superior de Tecnologías y Ciencias Aplicadas	ISTCA	18	535	2564	4.5	28
28	Puerto Rico	Universidad Central del Caribe	UCEC	8	642	5527	10.5	56
29	Cuba	Instituto Nacional de Ciencias Agrícolas	INCA	33	1171	2937	13.2	40
30	Guatemala	Universidad de San Carlos de Guatemala	USCG	27	441	2888	6.7	43

Conclusión

- La realización de este algoritmo permite tener una herramienta que puede ser utilizada por otras universidades para extraer datos y evaluar indicadores científicos de sus investigadores.
- La limitante de algoritmo es que es bloqueado por Google Académico si es usado en largos periodos de tiempo al ser considerado un robot de extracción de datos.
- Uno de los objetivos a largo plazo es hacer análisis de los perfiles de las revistas científicas en Google Académico.

MUCHAS GRACIAS

www.revistas.utp.ac.pa

www.ridda2.utp.ac.pa

www.investigadores.utp.ac.pa

Twitter @utprepositorio

Facebook /utprepositorio

utp.repositorio@utp.ac.pa